ORIGINAL RESEARCH

RETRACTED ARTICLE: HS3ST2 and Its Related Molecules as Potential Biomarkers for Predicting Lymph Node Metastasis in Patients with Colorectal Cancer

Guangyu Gao [iD]*
Xinya Shi
Jiaofeng Shen*

Department of Oncology, The Second Affiliated Hospital of Soochow University, Suzhou, 215004, Jiangsu, People's Republic of China

*These authors contributed equally to this work

Background: Lymph node metastasis is a major cause of cancer-related death in patients with colorectal cancer (CRC), but current strategies are limited to predicting this clinical behavior. Our study aims to establish a lymph node metastasis prediction model based on miRNA and mRNA to improve the accuracy of prediction.

Methods: GSE56350, GSE70574, and GSE95109 were downloaded from the Gene Expression Omnibus (GEO) database and 569 colorectal cancer statistics were also downloaded from The Cancer Genome Atlas (TCGA) database. Differentially expressed miRNAs were calculated by using R software. Besides, gene ontology and enriched pathway analysis of target mRNAs were analyzed by using FunRich. Furthermore, the mRNA–miRNA network was constructed using Cytoscape software. Gene expression level was also detected by performing qRT-PCR (quantitative real-time PCR) in colorectal cancer and lymph node tissues.

Results: In total, differentially expressed miRNAs were selected, and 34 mRNAs were identified after filtering. The research of KEGG indicated that mRNAs are enriched in many cancer pathways. Differentially expressed miRNAs were most enriched in the cytoplasm, nucleoside, transcription factor activity, and RNA binding. KEGG pathway analysis of these target genes was mainly enriched in 5 pathways including fatty acid elongation, MAPK signaling pathway, autophagy, signaling pathways regulating pluripotency of stem cells, and Th17 cell differentiation. The results of qRT-PCR indicated that hsa-miR-100 and hsa-miR-99a were differentially expressed in lymph node metastatic colorectal cancer tissues and lymph node non-metastasis tissues which all target HS3ST2. Besides, we also found they have a significant difference in colorectal cancer tissues compared with normal tissues.

Conclusion: By using microarray and bioinformatics analyses, differentially expressed miRNAs were identified and a complete gene network was constructed. To our knowledge, HS3ST2 and related molecules including hsa-miR-100 and hsa-miR-99a were firstly identified as potential biomarkers in the development of lymph node metastatic colorectal cancer.

Keywords: microRNA, colorectal cancer, lymph node metastasis, prognostic signature

Introduction

Colorectal cancer (CRC) is a serious health threat worldwide. Unfortunately, ~25% to 40% will develop a tumor recurrence despite a potentially curative operation. Compared with the early stage, the response and overall survival of patients with advanced colorectal cancer are still very poor.[1] Surgical tumor resection is still the cornerstone of the treatment of local advanced colorectal cancer. There is no curable treatment for metastatic tumors that cannot be surgically removed, as well as those with poor chemotherapy and radiotherapy effects.[2] At present, AJCC's TNM staging

Correspondence: Jiaofeng Shen
Department of Oncology, The Second Affiliated Hospital of Soochow University, 1055, Sanxiang Road, Suzhou, 215004, Jiangsu, People's Republic of China
Email jfshenzz@163.com

system has limited value in predicting recurrence.[3,4] Therefore, it is urgent to find out the factors influencing the recurrence of colorectal cancer, to promote the prognosis evaluation and individualized treatment.

MicroRNAs (miRNAs) are small, non-coding RNAs, having the function of regulating after gene transcription.[5] According to previous studies, miRNAs can regulate many target genes or one type of miRNAs can be regulated by many genes.[6] In particular, Thomas et al found that some miRNAs can improve the therapeutic effect by improving the drug sensitivity of cancer cells.[7] Xuan et al found that miR-374a, miR-92a, and miR-106a increase drug resistance and promote growth and metastasis of lung cancer.[8] Kania et al also reported that miR-9-3p and miR-9-5p decrease DNA topoisomerase IIα expression levels in acquired resistance to etoposide and may act as biomarkers of responsiveness to TOP2-targeted therapy.[9] However, the mechanisms of miRNAs in the adenoma transformation to adenocarcinoma remain unknown.

In this study, microarray data from the GEO database and colorectal cancer sample data in the TCGA database promoted the research of differently expressed miRNAs in lymph node non-metastasis cancer tissues and lymph node metastatic cancer tissues. The functions of the target mRNAs were evaluated by using GO annotation, KEGG, PPI network, and the association between identified miRNAs and the overall survival of patients with cancer. Our goal is to identify genes associated with lymph node metastasis colorectal cancer and subsequent pathways.

Materials and Methods
Microarray Data

The GEO (https://www.ncbi.nlm.nih.gov/geo) database is a public functional genomics data set to help users download statistics and import gene expression. In our research, Gene expression profile data (GSE56350, GSE70574, and GSE95109) were obtained from GEO.

GSE56350, including 8 primary colorectal cancer tissues derived from stage II–III colorectal cancer patients with (n = 20) or without (n = 15) lymph node metastasis. MicroRNA expression profiling analysis of these samples was performed on Agilent-021827 Human miRNA Microarray [miRNA_107_Sep09_2_105]. Dataset GSE70574, 16 T1-stage CRCs (7 LNM-positive and 9 LNM- negative tumors) were processed by Agilent-031181 Unrestricted_Human_miRNA_V16.0_Microarray 030840 (Feature Number version). GSE95109, 13 patients

in lymph node-negative subgroup and 9 patients in lymph node-positive subgroup. The mRNA profiles of all 22 patients were analyzed by Agilent microarray technology to explore the different expressions between the lymph node-negative subgroup and lymph node-positive subgroup.

Differently Expressed miRNAs and mRNAs Analysis

The raw count data were firstly normalized with transcripts per million (TPM) method and underwent a log2 transformation. R software was used to compare two groups of tissues using the Limma version 3.36.2 R package. Besides, |log2FC|≥1 and <0.05 were used as a cut-off criterion and a significant statistical difference would be considered if the statistics met our standards.[10]

Functional and Pathway Enrichment Analysis

Funrich (http://www.funrich.org) is a publicly accessible software with the ability to identify the enriched transcription factors and perform Gene Ontology (GO) functional analysis of upload differentially expressed miRNAs. In this study, transcription factors enrichment analysis was carried out and 10 transcription factors that may regulate differentially expressed miRNAs were identified. Besides, we also used this software to obtain target genes of differentially expressed miRNAs. KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis was performed by using Cytoscape software. ClueGO is a Cytoscape plug-in that visualizes the non-redundant biological terms for large clusters of genes in a functionally grouped network. The network graph of ClueGO is created based on kappa statistics. Each node in the graph represents a term. The connection between nodes reflects the correlation between terms, while the color of nodes reflects the enrichment and classification of the node (which functional group it belongs to).

PPI Network and Clustered Subnetworks' Construction

The exploration of protein interactions helps to reveal the underlying pathological mechanism of colorectal cancer. In this study, we used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (https://string-db.org/) to construct a protein–protein interaction

network. By uploading all the target genes, PPI network including number and size of nodes was constructed.

Prediction of miRNA Target Genes and miRNA-mRNA Regulatory Network

A previous study suggested that the function of miRNAs lies in the regulation of target genes. Therefore, the prediction of target genes is particularly important. It can indirectly understand the biological function and enrichment pathway of miRNAs. The miRNA enrichment function in FunRich was used to achieve miRNA targeting prediction. Besides, GSE95109 was downloaded from the GEO database. By combining the results from FunRich and difference analysis of GSE95109, screened genes were identified and the miRNA-mRNA regulatory network was built by using Cytoscape software.

Construction of a Prognostic Signature Model

To find out the influence of differential expression of miRNAs on the prognosis of CRC patients, the univariate and multivariate Cox proportional risk regression analysis was carried out for different expression of miRNAs, the miRNAs related to CRC prognosis were selected, and the risk linear model based on TCGA data set was established. Five hundred and sixty-one CRC patients downloaded from TCGA were randomly divided into training groups and test groups. We built a predictive feature model in the training group and tested its validity in the test group. First, the training group was

analyzed by univariate Cox regression analysis to select the prognoses related to DEGs. In the further functional analysis and development of potential risk characteristics, the least absolute shrinkage and selection operator (LASSO) method was used to regress the high-dimensional prediction factors.[11,12] The "glmnet" packet in R was used to calculate the coefficient and partial likelihood deviation.[13] Through multivariate Cox regression analysis, we further studied these miRNAs to identify significant targets and build a risk linear model. To better understand the relationship between the selected miRNAs and the prognosis of colorectal cancer patients, we constructed a risk prediction model. By using "survival ROC" package, we calculated the AUC (Area Under Curve) of 3 years and 5 year dependent ROC (Receiver Operating Characteristic) curve to assess the predictive power of identified miRNAs.

Verification of miRNA Expression

To validate the data from miRseq, 2 miRNAs (hsa-miR-100 and hsa-miR-99a), which were found to be important in the above bioinformatics analysis, were selected and validated with real-time qPCR. Specific divergent primers for each miRNA were designed according to the sequences of the linear transcripts, and all divergent primers are shown in Table 1. Total RNA was extracted from 20 pairs of lymph node non-metastasis cancer tissues and lymph node metastatic cancer tissues, and digested using RNase R. cDNA was synthesized using the EasyScript One-Step gDNA Removal and cDNA

Table 1 The Primers for Quantitative Real-Time Polymerase Chain Reaction, with β-Actin and U6 Used as the Control

Primer Sets Name	Reverse Transcriptase Primers (5′ to 3′)	RT-qPCR Primers (5′ to 3′)
hsa-mir-100	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCACAAGT	F: CAGTGCAGGGTCCGAGGTAT R: CGCGAACCCGTAGATCCGAA
hsa-mir-99a	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCACAAGA	F: CAGTGCAGGGTCCGAGGTAT R: CGTATAACCCGTAGATCCGAT
U6	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAAAATA	F: AGAGAAGATTAGCATGGCCCCTG R: ATCCAGTGCAGGGTCCGAGG
HS3ST		F: GGCTGGATTGGTACAGGAGC R: CCACGATCAGCTTGGTGTCT
β-actin		F: CTCCATCCTGGCCTCGCTGT R: GCTGTCACCTTCACCGTTCC

Synthesis SuperMix (Transgen Biotech, China). Real-time PCR was performed on the QuantStudio 5 Real-Time PCR System (Thermo Fisher) using qPCR SYBR Green master mix (CloudSeq). The expression levels of miRNAs were normalized to U6 (internal standard control) and calculated using the 2−ΔΔCt method. All experiments were performed in triplicate.

Verification of mRNA Expression

The common potential target gene HS3ST2 of 2 miRNAs (hsa-miR-100 and hsa-miR-99a) was obtained through bioinformatics analysis, and then real-time qPCR verification was conducted. A specific divergent primer was designed according to the sequences of the linear transcripts, and all divergent primers are shown in Table 2. Total RNA was extracted from 20 pairs of lymph node non-metastasis cancer tissues lymph node metastatic cancer tissues and digested using RNase R. cDNA was synthesized using the EasyScript One-Step gDNA Removal and cDNA Synthesis SuperMix (Transgen Biotech, China). Real-time PCR was performed on the QuantStudio 5 Real-Time PCR System (Thermo Fisher) using qPCR SYBR Green master mix (CloudSeq). The expression levels of mRNAs were normalized to β-actin (internal standard control) and calculated using the 2−ΔΔCt method. All experiments were performed in triplicate.

Hematoxylin and Eosin (H&E) Staining and Analysis

Fresh colorectal carcinoma and lymph node tissues were fixed in 10% formalin and embedded in paraffin before sectioning and staining. Tissue sections 4 μm thick were deparaffinized in xylene and rehydrated in ethanol series. H&E staining was performed according to standard protocols.

Results

Identification of the miRNAs Between Lymph Node Non-Metastasis Tissues and Lymph Node Metastasis Tissues

R software was used to research the gene expression profiles from the GSE56350, GSE70574, and GSE95109. According to the cut-off criteria ($P<0.05$ and $|log2FC|\geq 1$), GSE56350 and GSE70574 were selected from GEO datasets. The volcano and heatmap of the two datasets were, respectively, shown in Figure 1A–D. A total of 5 common differentially expressed miRNAs were screened from GSE56350 and GSE70574 datasets (Figure 2). The basic information of the 5 miRNAs is listed in Table 1. Besides, the differentially expressed mRNAs between the lymph node-negative subgroup and lymph node-positive subgroup were obtained from GSE95109 (Figure 1E and F) including 34 mRNAs (29 upregulated and 5 downregulated).

Screening of Potential Transcription Factors and Enrichment Analysis

For the reason that transcription factors are crucial in miRNA, FunRich is used to research the top 10 enriched transcription, namely SP1, TEAD1, TCF3, SOX1, HNF4A, TFAP4, KLF7, NHLH1, HENMT1, and CREB1. To learn more about the function of these miRNAs, we uploaded them into FunRich to perform GO analysis. The result showed that differentially expressed miRNAs were most enriched in the cytoplasm, nucleoside, transcription factor activity, and RNA binding (Figure 3A and B). KEGG pathway analysis of these target genes was performed by using Cytoscape and ClueGO. These selected mRNAs were mainly enriched in 5 pathways including Fatty acid elongation, MAPK signaling pathway, Autophagy, Signaling pathways

Table 2 Differentially Expressed (DE) miRNAs Identified from GSE56350 and GSE70574 Dataset

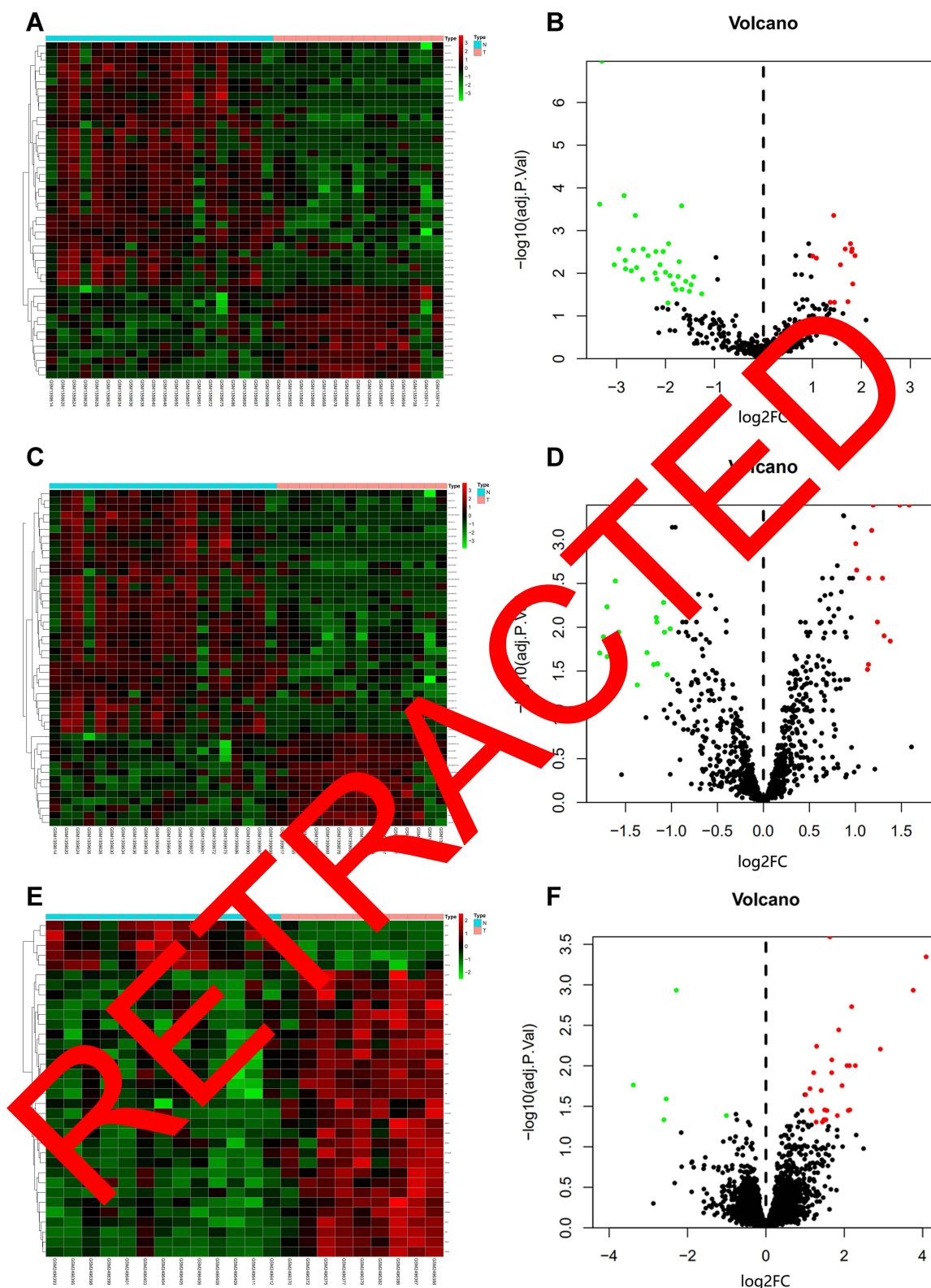| ID	logFC	AveExpr	t	P.Value	Adj. P.Val
hsa-miR-99a	−3.2998	5.878924	−9.07224	1.47E-10	1.08E-07
hsa-miR-100	−2.35603	6.687555	−4.37844	0.00011	0.003868
hsa-miR-125b	−2.95247	5.767905	−4.71345	4.12E-05	0.002706
hsa-miR-143	−3.34466	5.496491	−5.96943	9.83E-07	0.000242
hsa-miR-375	−1.95009	7.292945	−3.1063	0.003834	0.049641

Figure 1 Volcano plot and heat map of GSE56350, GSE70754, and GSE95109. (**A**) Unsupervised clustering analysis of differentially expressed (DE) miRNAs in GSE56350. Red dots indicate significantly up-regulated miRNAs, green dots indicate significantly down-regulated miRNAs. (**B**) Volcano plots of miRNAs in GSE56350. Red dots indicate up-regulated DEmiRNAs, green dots indicate down-regulated DEmiRNAs, black dots indicate non-differentially expressed miRNAs. (**C**) Unsupervised clustering analysis of DE miRNAs in GSE70754. Red dots indicate significantly up-regulated miRNAs, green dots indicate significantly down-regulated miRNAs. (**D**) Volcano plots of miRNAs in GSE70754. Red dots indicate up-regulated DEmiRNAs, green dots indicate down-regulated DEmiRNAs, black dots indicate non-differentially expressed miRNAs. (**E**) Unsupervised clustering analysis of the DE miRNAs in GSE95109. Red dots indicate significantly up-regulated miRNAs, green dots indicate significantly down-regulated miRNAs. (**F**) Volcano plots of miRNAs in GSE95109. Red dots indicate up-regulated DEmiRNAs, green dots indicate down-regulated DEmiRNAs, black dots indicate non-differentially expressed miRNAs.
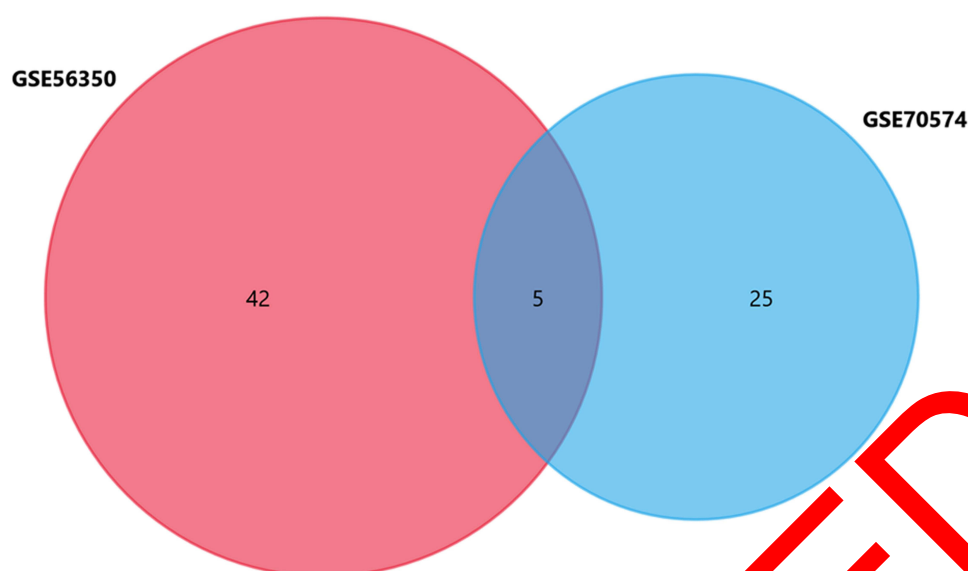
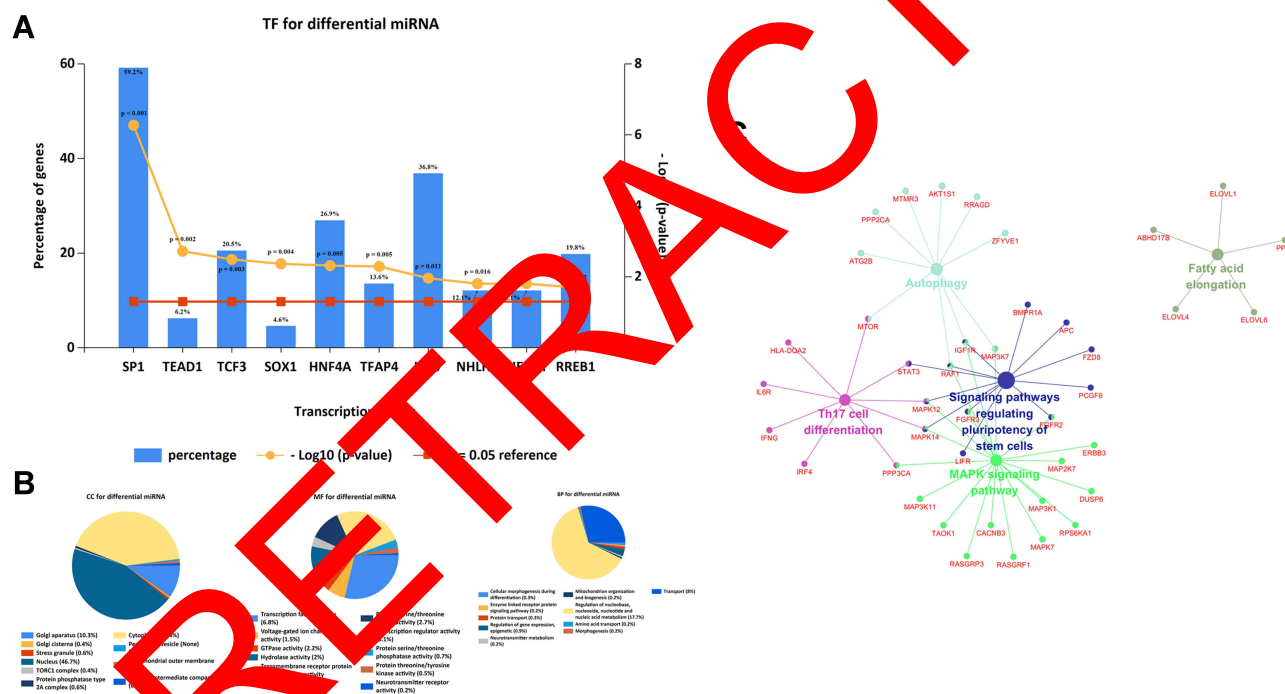**Figure 2** Venn Diagram of GSE48074 and GSE70574.

Figure 3 Screening of potential transcription factors and target genes of differentially expressed miRNAs. (**A**) Identification of the potential transcription factors of DEMs by FunRich software. (**B**) The top 10 of biological process, cellular component, and molecular function of the target genes of miRNAs. (**C**) KEGG pathway enriched by potential target mRNAs.

regulating pluripotency of stem cells, and Th17 cell differentiation (Figure 3C). Besides, the construction of the protein–protein interaction (PPI) network is shown in Figure 4. This network included 598 nodes and 1004 edges, under the conditions that the comprehensive Gt score >0.7.

miRNA-mRNA Regulatory Network

By using FunRich software, 598 potential target genes of 5 DE-miRNAs were obtained and only 1 of them differentially expressed in GSE95109. The Venn Diagram showed the results we screened (Figure 5A). To show the composition and relationship of target genes more intuitively, a
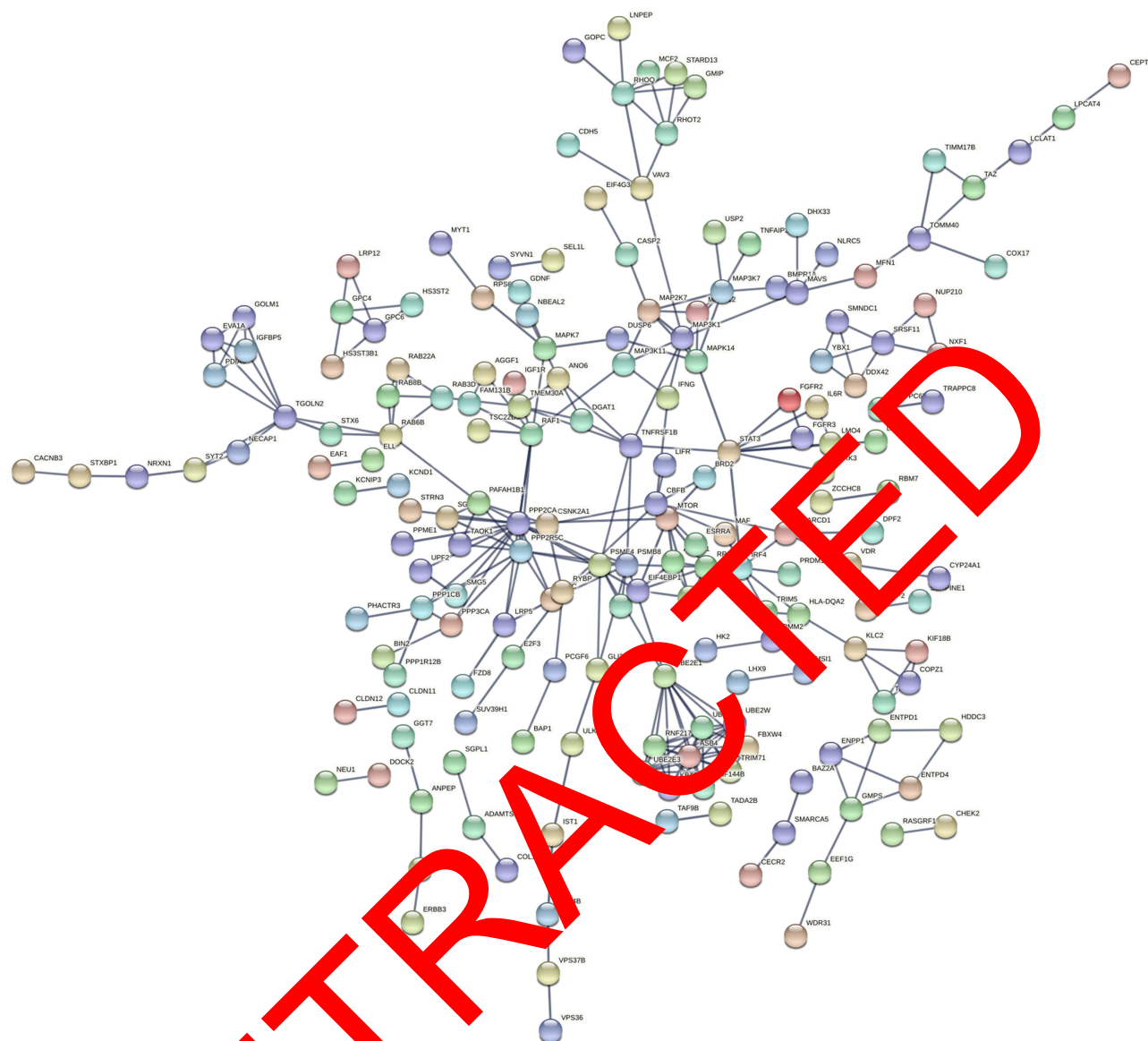
Figure 4 The PPI network of the target genes of the identified DEmRNAs.

complete network of target genes was constructed by using Cytoscape. Finally, 2 essential miRNA-mRNA pairs including hsa-miR-100, hsa-miR-99a and common target gene HS3ST2 were identified which implied the crucial effect of lymph node metastasis colorectal cancer (Figure 5B).

Construction of Prognostic Risk Model and Predictability Assessment

To identify the best prognostic miRNAs, the LASSO Cox regression algorithm was applied for 20 prognostic-related miRNAs. Nine miRNAs were selected to build the risk signature based on the minimum standard (Figure 6A).

Then, multivariate Cox proportional risk regression analysis was carried out for 9 candidate prognostic miRNAs to assess their independent prognostic values. According to Cox model, 7 candidate miRNAs (hsa-miR-125a-5p, hsa-miR-377, hsa-miR-100, hsa-miR-455-3p, hsa-miR-126, hsa-miR-199a, and hsa-miR-99a) were selected as independent significant prognostic factors. Seven prognostic miRNAs were then combined to build a model to predict patient outcomes. The AUC of 3 years survival for the 7-miRNA signature achieved 0.809 and the AUC of 5 years survival achieved 0.981, which proved that the model has good performance in predicting the survival risk of colorectal cancer patients (Figure 6B). According to this risk
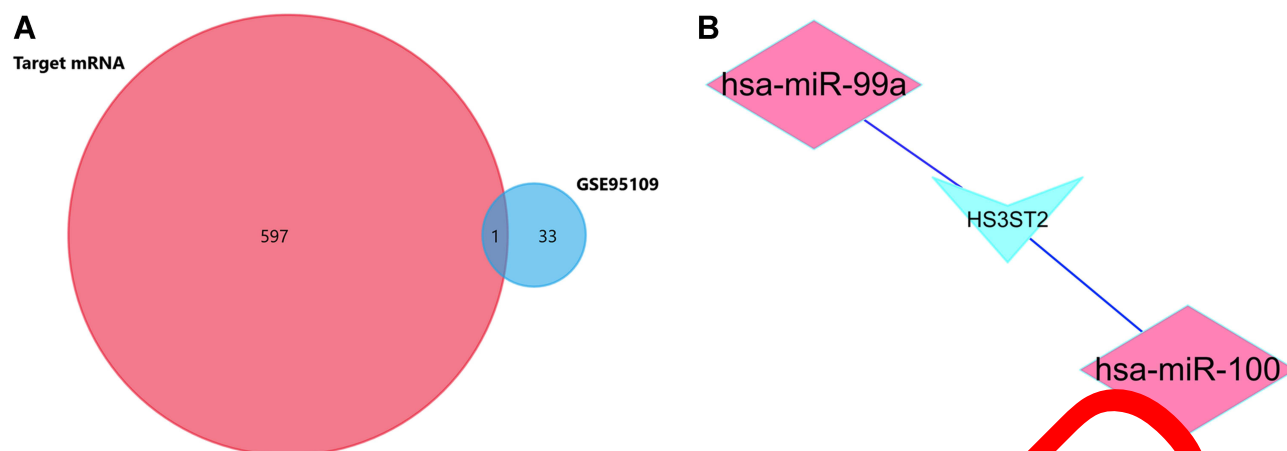
Figure 5 The miRNA-mRNA network of hsa-miR-100, hsa-miR-99a, and HS3ST2 in CRC tissues and CRCLN tissues. (**A**) Venn diagram of target mRNAs and GSE95109. (**B**) Identified target mRNAs and miRNA-mRNA regulatory network.

model, patients were divided into high- and low-risk groups. The results show that this model can well predict the clinical outcomes of patients. We also analyzed the risk score, survival status, and distribution of 7 miRNAs' expressions in each patient (Figure 6C and D).

Hsa-miR-100, Hsa-miR-99a, and HS3ST2 Were Significantly Differentially Expressed in CRC Tissues and Their Corresponding Normal-Appearing Tissues

By using FunRich software, we also found that hsa-miR-100 and hsa-miR-99a had high reliability, which all target HS3ST2. To further explore the expression relationship of hsa-miR-99a, hsa-miR-100, and HS3ST2 in CRC tissues and their corresponding normal tissues, R software was used to research the gene expression profiles from the GSE126093, GSE108153, GSE146587, and GSE110224. According to the cut-off criteria (p<0.05 and |log2FC|≥1), GSE126093, GSE108153, GSE146587, and GSE110224 were selected from GEO datasets. hsa-miR-99a and hsa-miR-100 were screened from the GSE108153 and GSE126093 dataset, and the expression level of the 2 miRNAs is shown in Figure 7A–D. Besides, the results showed that hsa-miR-99a and hsa-miR-100 were down-regulated in both GSE108153 and GSE126093. As for HS3ST2, it was also screened from the GSE146587 and GSE110224 dataset, and the expression level of HS3ST2 is shown in Figure 7E and F. The results indicated that HS3ST2 was up-regulated in both GSE108153 and GSE126093.

Hematoxylin and Eosin (HE) Staining and Validation of the Expression with qRT-PCR

By staining, the comparison of HE morphology between the four groups is shown in Figure 8A–D. To further assess the expression of hsa-miR-100, hsa-miR-99a, and HS3ST2, a total of 20 pairs of lymph node non-metastasis cancer tissues (CRC) and lymph node metastatic cancer tissues (CRCLN) were enrolled as a validation cohort. The qRT-PCR method was used to confirm the differential expression levels from participant's samples. Consistent with the microarray data, hsa-miR-100 and hsa-miR-99a were significantly downregulated (Figure 8E and F) and HS3ST2 was upregulated (Figure 8G) between 20 pairs of lymph node non-metastasis cancer tissues and lymph node metastatic cancer tissues which indicated that hsa-miR-100, hsa-miR-99a, and HS3ST2 could be the candidate biomarkers for CRC and CRCLN.

Discussion

Colorectal cancer can occur through the progression of adenomas, which is the result of epithelial cell genetic and epigenetic events. Some microarray articles have identified gene expression profiles in adenomas and cancers.[14–16] In this present research, gene expression data of GSE56450, GSE70574, and GSE95109 were searched from the GEO. Five hundred and sixty-nine colorectal cancer statistics were searched from the TCGA database. We identified 6 cancer-related gene expression patterns, suggesting that these 5 miRNAs play a role in promoting the development of colorectal cancer. To learn more about
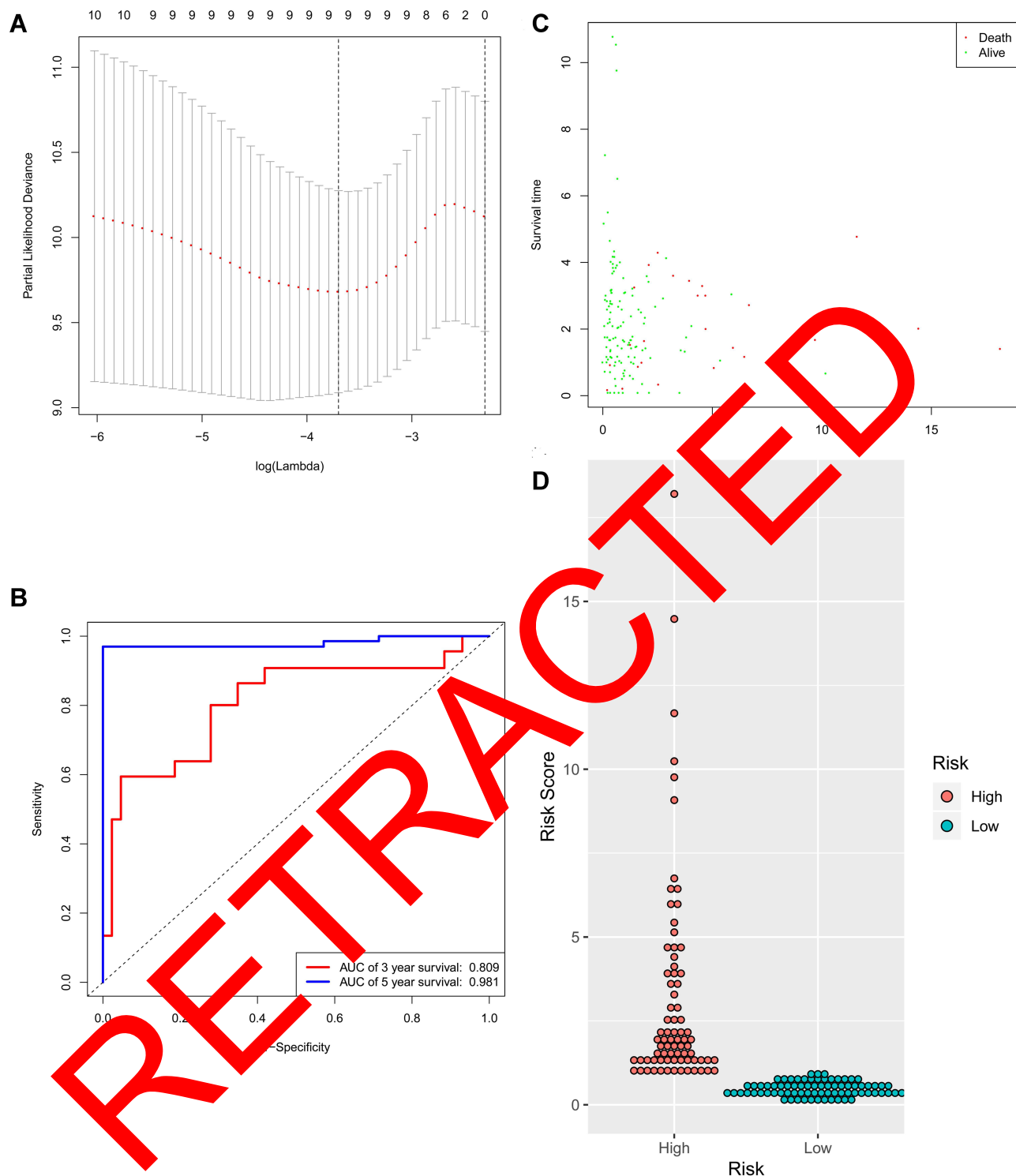
Figure 6 Identification of the signature significantly associated with the survival of patients with CRC in the training group. (**A**) LASSO Cox regression algorithm was used to reduce the scope. (**B**) Time-dependent ROC curves analysis. (**C** and **D**) Risk score distribution and survival status for patients in high- and low-risk groups by the signature (*p*<0.05).
Abbreviations: LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic.

the regulatory mechanism of the 5 miRNAs in colorectal cancer, we used FunRich for further study. Function annotation indicated that these miRNAs were primarily related to the cytoplasm, nucleoside, transcription factor activity, and RNA binding. This is consistent with the recognition that transcription factor activity and RNA binding are the
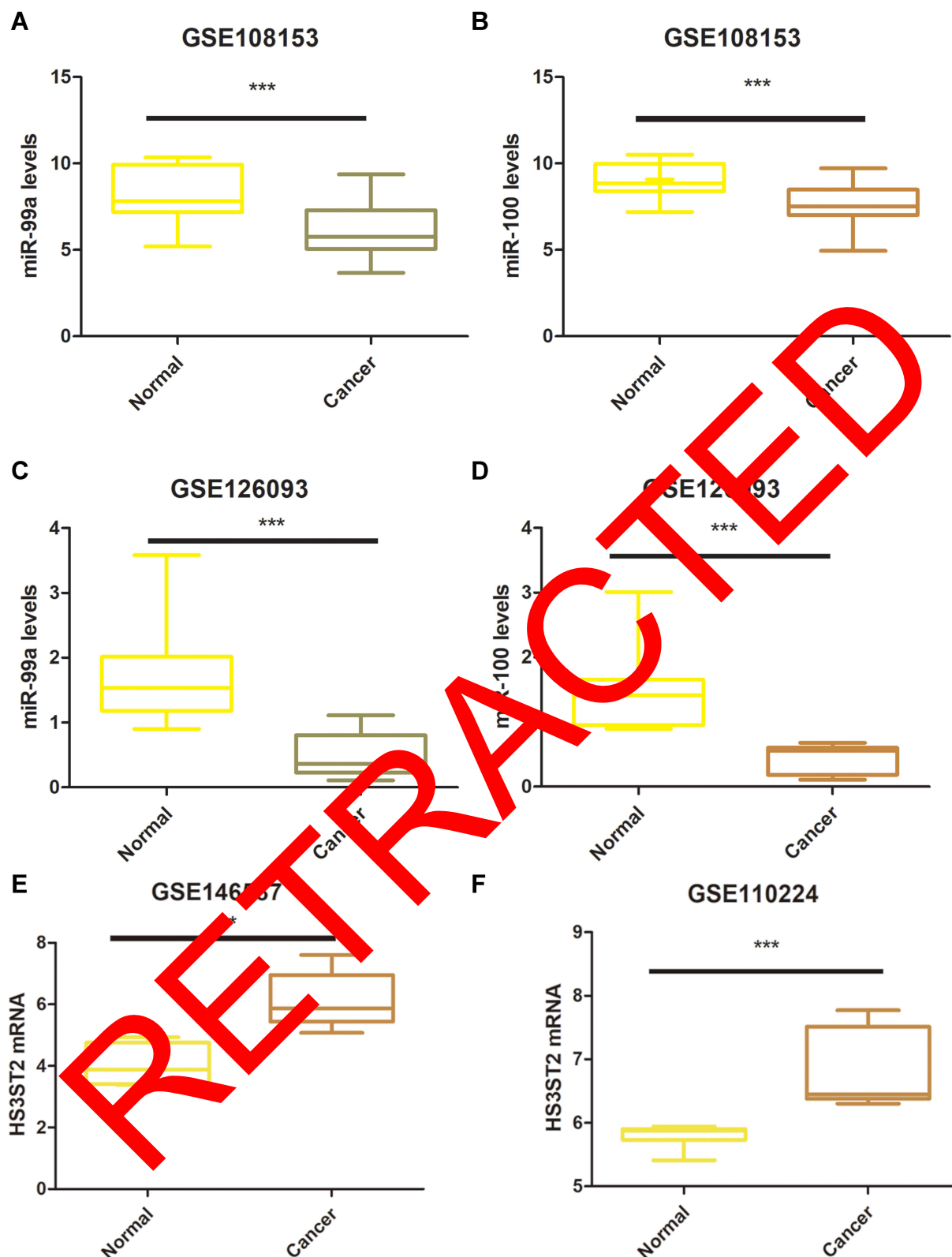
Figure 7 The differential expression of hsa-miR-100 and hsa-miR-99a in CRC tissues and their corresponding normal-appearing tissues. (**A**) Validation of hsa-miR-99a in GSE108153 (***$p < 0.001$). (**B**) Validation of hsa-miR-100 in GSE108153 (***$p < 0.001$). (**C**) Validation of hsa-miR-99a in GSE126093 (***$p < 0.001$). (**D**) Validation of hsa-miR-100 in GSE126093 (***$p < 0.001$). (**E**) Validation of HS3ST2 in GSE146587 (**$p < 0.01$). (**F**) Validation of HS3ST2 in GSE110224 (***$p < 0.001$).
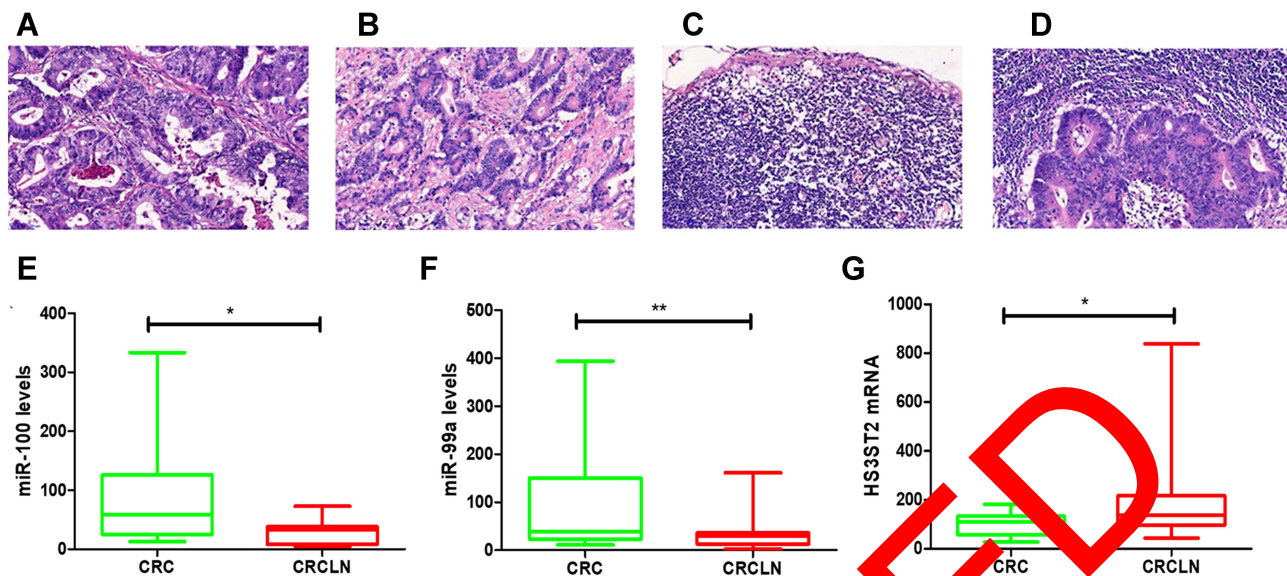
Figure 8 The expression level of hsa-miR-100, hsa-miR-99a, and HS3ST2 in CRC tissues and CRCLN tissue. (**A**–**D**) Hematoxylin and eosin (H&E) staining of CRC lymph node metastatic tissue and CRC lymph node nonmetastatic tissue (scale bar: 200×). (**E**) Validation of hsa-miR-100 (*$p < 0.05$). (**F**) Validation of hsa-miR-99a (**$p < 0.01$). (**G**) Validation of HS3ST2 (*$p < 0.05$).

main causes of tumor occurrence and development.[17,18] The cytoplasm in cancer cells is essentially different from that in normal cells.[19] KEGG pathway analysis indicated that these identified genes were mainly enriched in 5 pathways including Fatty acid elongation, MAPK signaling pathway, Autophagy, Signaling pathways regulating pluripotency of stem cells, and Th17 cell differentiation. Several studies on the MAPK signaling pathway have shown that it participates in a diverse array of important cellular processes, including the survival, proliferation, differentiation, and activation of different cell types.[20,21] As for the Th17 cell differentiation, a previous study provides evidence that the miR-146a acts as an important molecular brake that blocks the autocrine IL-6- and IL-21-induced Th17 differentiation pathways in autoreactive CD4 T cells, highlighting its potential as a therapeutic target for treating autoimmune diseases.[22] Also, recent researches showed that autophagy influences the interaction between the tumor and the host by promoting stress adaptation and suppressing activation of innate and adaptive immune responses. Besides, it can promote cancer by suppressing p53 and preventing energy crisis, cell death, senescence, and an anti-tumor immune response.[23] However, the regulation mechanism of these pathways in colorectal cancer has not been reported yet.

MiRNA-mRNA regulatory network was built based on FunRich and Cytoscape. Five differently expressed miRNAs (hsa-miR-100, hsa-miR-375, hsa-miR-125b,

hsa-miR-143, and hsa-miR-99a) and 1 potential gene were identified by combining two screening results. Among them, some miRNAs have been shown to affect tumor proliferation, migration, and prognosis. For example, a previous study reported down-regulation of hsa-miR-100 in Non-small cell lung cancer (NSCLC) altered the expression of GRP78 and spliced XBP1 level within UPR pathway. It may affect ER stress and lung cancer and be used as a diagnostic biomarker of activated UPR in NSCLC.[24] As for hsa-miR-99a, a previous study reported low expression of miR-99a significantly predicts poor prognosis in head and neck squamous cell carcinoma and regulates cancer cell migration and invasion.[25] Xu et al also demonstrated that miR-99a inhibited the migratory and invasive abilities by regulating the expression of insulin-like growth factor 1 receptor and the miR-99a/IGF1R axis may provide novel insight into the pathogenesis of gastric cancer.[26] As for hsa-miR-125b, a previous study reported overexpression of hsa-miR-125b significantly enhanced the ability of cell proliferation, migration, and invasion in Head and Neck Squamous Cell Carcinoma, with upregulation of C-C Chemokine Receptor Type 7.[27] It also activates p53 and Induces Apoptosis in Lung Cancer Cells.[28] Besides, hsa-miR-125b may act as a cancer suppressor by regulating the BRCA1 signaling, and reintroduction of hsa-miR-125b analogs could be a potential adjunct treatment for advanced/chemoresistant breast cancers.[29]

Besides, a 7-miRNAs signature (hsa-miR-125a-5p, hsa-miR-377, hsa-miR-100, hsa-miR-455-3p, hsa-miR-126, hsa-miR-199a, and hsa-miR-99a) of 569 patients with colorectal cancer were constructed by using single variable and multivariate Cox and risk scoring methods. CRC patients were divided into a high-risk group and a low-risk group. The results showed that the prognosis predictive performance of the 7-miRNAs signature was good in the TCGA CRC cohort. Besides, the 7-miRNAs risk was an independent prognostic factor of CRC and patients in the high-risk group showed significantly poorer survival than patients in the low-risk group. ROC demonstrated that the nomogram combining the 7-miRNAs signature performed the best in predicting short-term survival (3-year and 5-year) for patients with CRC. All these results indicated that the risk model developed from the 7 genes could be a useful indicator for CRC survival.

Through the combination of the GEO and TCGA analysis results, 2 miRNAs (hsa-miR-100 and hsa-miR-99a) were identified for further research. We also found that hsa-miR-100 and hsa-miR-99a had strong effects, which all target HS3ST2. By using qRT-PCR and IHC, we also found that they have different expressions in lymph node non-metastasis tissues and lymph node metastasis tissues. The HS3ST2 gene, an enzyme mediating the 3-O-sulfation of heparan sulfate (HS), was discovered and cloned in 2001.[30] It is present in all cell types and tissues and functionally interacts with growth factors, tyrosine kinase receptors, matrix metalloproteinases (MMPs), and extracellular matrix (ECM) proteins to modulate cell adhesion, proliferation, and motility.[31] In breast cancer, colorectal cancer, lung cancer, cervical cancer, pancreatic cancer, and recurrent prostate cancer, HS3ST2 is silenced due to hypermethylation, suggesting a potentially important pathogenic role.[32] A previous study confirmed that abnormal methylation levels of the HS3ST2 gene is important in endometrial hyperplasia and carcinogenesis.[33] Zuo et al also reported that the expression of HS3ST2 was upregulated in the methylation-positive cervical tissues than in the methylation-negative cervical tissues. Besides, it may play important roles in HPV-induced cervical cancer, and that patients with specific hypermethylated genes may have a greater risk of progressing to invasive cervical cancer.[33] The prognostic significance of HS3ST2 mRNA expression in several cancer types has been evaluated.[34,35] HS3ST2 protein expression could be used as a favorable prognostic tissue biomarker in patients with primary advanced-stage lung cancer.[35] For gastric cancer, statistical analyses using a chi-squared test showed that there is a statistically significant difference in methylation level of HS3ST2 genes between gastric cancer and uncancerous patients and it may act as novel cancer-related molecular mechanisms in the detection of new treatment strategies.[36]

Nowadays, with the development of precision medicine, more and more attention has been paid to individual differences in treatment methods. Therefore, it is necessary to find new biomarkers and treatment methods for patients. Our findings approved that many differentially expressed mRNAs and miRNAs involved in the lymph node metastasis of colorectal cancer by certain signaling pathways and had prognostic value. Since all our data were mostly achieved from the GEO database and TCGA by bioinformatics tools, as well as the limited number of relevant samples, more data analysis and clinical experiments should be further performed for developing these potential biomarkers for predicting the recurrence of CRC.

Conclusion

Our study concluded certain mechanisms of the development of colorectal cancer. Plenty of differentially expressed mRNAs and miRNAs were identified between lymph node non-metastasis tissues and lymph node metastasis tissues by using bioinformatics methods. Also, hsa-miR-100, hsa-miR-99a, and HS3ST2 were identified as potential biomarkers for predicting the recurrence of colorectal cancer.

Abbreviations

GEO, Gene Expression Omnibus; KEGG, Kyoto Encyclopedia of Genes and Genomes; miRNA, microRNAs; TCGA, The Cancer Genome Atlas; ROC, receiver operating characteristic; AUC, area under the curve.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the Yulan Gu on reasonable request.

Author Contributions

Guangyu Gao and Xinya Shi contributed equally to this work. All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval

of the version to be published; and agree to be accountable for all aspects of the work.

Disclosure

The authors have declared that no competing interest exists.

References

1. Steele SR, Chang GJ, Hendren S, et al.; Practice Guidelines Committee of the American Society of C. Practice guideline for the surveillance of patients after curative treatment of colon and rectal cancer. *Dis Colon Rectum*. 2015;58(8):713–725. doi:10.1097/DCR.0000000000000410

2. Vogel JD, Eskicioglu C, Weiser MR, Feingold DL, Steele SR. The American Society of Colon and Rectal Surgeons clinical practice guidelines for the treatment of colon cancer. *Dis Colon Rectum*. 2017;60(10):999–1017. doi:10.1097/DCR.0000000000000926

3. Balch C, Ramapuram JB, Tiwari AK. The epigenomics of embryonic pathway signaling in colorectal cancer. *Front Pharmacol*. 2017;8:267. doi:10.3389/fphar.2017.00267

4. Dienstmann R, Mason MJ, Sinicrope FA, et al. Prediction of overall survival in stage II and III colon cancer beyond TNM system: a retrospective, pooled biomarker study. *Ann Oncol*. 2017;28 (5):1023–1031. doi:10.1093/annonc/mdx052

5. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015;21(11):1350–1356. doi:10.1038/nm.3967

6. Ren P, Gong F, Zhang Y, Jiang J, Zhang H. MicroRNA-92a promotes growth, metastasis, and chemoresistance in non-small cell lung cancer cells by targeting PTEN. *Tumour Biol*. 2016;37(3):3215–3225. doi:10.1007/s13277-015-4150-3

7. Thomas SK, Wang F, Meng F, et al. Implications of MicroRNA in the treatment of gefitinib-resistant non-small cell lung cancer. *Int J Mol Sci*. 2016;17(2):237. doi:10.3390/ijms17020237

8. Xuan M, Liang AL, Liu YJ. Research progress on the relationship between lung cancer drug-resistance and microRNAs. *J Cancer*. 2019;10(27):6865–6875. doi:10.7150/jca.31952

9. Kania EE, Carvajal-Moreno J, Hernandez VA, et al. hsa-miR-9-3p and hsa-miR-9-5p as post-transcriptional modulators of DNA topoisomerase iialpha in human leukemia K562 cells with acquired resistance to etoposide. *Mol Pharmacol*. 2020;97(3):159–170. doi:10.1124/mol.119.118315

10. Li L, Wang G, Li N, Yu H, Si J, Wang J. Identification of key genes and pathways associated with obesity in children. *Exp Ther Med*. 2017;14(2):1065–1073. doi:10.3892/etm.2017.4597

11. Ikeda N, Nakajima Y, Tokuhara T, et al. Clinical significance of aminopeptidase N/CD13 expression in human pancreatic carcinoma. *Clin Cancer Res*. 2003;9(4):1503–1508.

12. Veenman CJ, Tax DM. LESS: a model-based classifier for sparse subspaces. *IEEE Trans Pattern Anal Mach Intell*. 2005;27(9):1496–1500. doi:10.1109/TPAMI.2005.182

13. Pollock BE, Storlie CB, Link MJ, Stafford SL, Garces YI, Foote RL. Comparative analysis of arteriovenous malformation grading scales in predicting outcomes after stereotactic radiosurgery. *J Neurosurg*. 2017;126(3):852–858. doi:10.3171/2015.11.JNS151300

14. Kitahara O, Furukawa Y, Tanaka T, et al. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res*. 2001;61(9):3544–3549.

15. Lechner S, Muller-Ladner U, Renke B, Scholmerich J, Ruschoff J, Kullmann F. Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia. *Gut*. 2003;52(8):1148–1153. doi:10.1136/gut.52.8.1148

16. Notterman DA, Alon U, Sierk AJ, Levine AJ. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res*. 2001;61 (7):3124–3130.

17. Perez R, Wu N, Klipfel AA, Beart RW Jr. A better cell cycle target for gene therapy of colorectal cancer: cyclin G. *J Gastrointest Surg*. 2003;7(7):884–889. doi:10.1007/s11605-003-0034-8

18. Tominaga O, Nita ME, Nagawa H, Fujii S, Tsuruo T, Muto T. Expressions of cell cycle regulators in human colorectal cancer cell lines. *Jpn J Cancer Res*. 1997;88(9):855–860. doi:10.1111/j.1349-7006.1997.tb00461.x

19. Djamgoz MB, Coombes RC, Schwab A. Ion transport and cancer: from initiation to metastasis. *Philos Trans R Soc Lond B Biol Sci*. 2014;369(1638):20130092. doi:10.1098/rstb.2013.0092

20. Sun Y, Liu WZ, Liu T, Feng X, Yang N, Zhou HF. Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. *J Recept Signal Transduct Res*. 2015;35(6):600–604. doi:10.3109/10799893.2015.1030412

21. Chen SX, Zhao F, Huang YF. [MAPK signaling pathway and erectile dysfunction]. *Zhonghua Nan Ke Xue*. 2018;24(5):442–446. Chinese.

22. Li B, Wang X, Choi IY, et al. miR-146a modulates autoreactive Th17 cell differentiation and regulates organ-specific autoimmunity. *J Clin Invest*. 2017;127(10):3702–3716. doi:10.1172/JCI94012

23. Amaravadi R, Kimmelman AC, White E. Recent insights into the function of autophagy in cancer. *Genes Dev*. 2016;30(17):1913–1930. doi:10.1101/gad.287524.116

24. Ahmadi A, Khansarinejad B, Hosseinkhani S, Ghanei M, Mowla SJ. miR-199a-5p and miR-495 target GRP78 within UPR pathway of lung cancer. *Gene*. 2017;620:15–22. doi:10.1016/j.gene.2017.04.032

25. Okada R, Koshizuka K, Yamada Y, et al. Regulation of oncogenic targets by miR-99a-3p (passenger strand of miR-99a-duplex) in head and neck squamous cell carcinoma. *Cells*. 2019;8(12):1535. doi:10.3390/cells8121535

26. Xu XL, Guo AX, Pan QY, Chang AL, Zhao CR. MiR-99a suppresses cell migration and invasion by regulating IGF1R in gastric cancer. *Eur Rev Med Pharmacol Sci*. 2019;23(17):7375–7382. doi:10.26355/eurrev_201909_18845

27. Jin S, Liu MD, Wu H, et al. Overexpression of hsa-miR-125a-5p enhances proliferation, migration and invasion of head and neck squamous cell carcinoma cell lines by upregulating C-C chemokine receptor type 7. *Oncol Lett*. 2018;15(6):9703–9710. doi:10.3892/ol.2018.8564

28. Jiang L, Chang J, Zhang Q, Sun L, Qiu X. MicroRNA hsa-miR-125a-3p activates p53 and induces apoptosis in lung cancer cells. *Cancer Invest*. 2013;31(8):538–544. doi:10.3109/07357907.2013.820314

29. Xu X, Lv YG, Yan CY, Yi J, Ling R. Enforced expression of hsa-miR-125a-3p in breast cancer cells potentiates docetaxel sensitivity via modulation of BRCA1 signaling. *Biochem Biophys Res Commun*. 2016;479(4):893–900. doi:10.1016/j.bbrc.2016.09.087

30. Afratis N, Gialeli C, Nikitovic D, et al. Glycosaminoglycans: key players in cancer cell biology and treatment. *FEBS J*. 2012;279 (7):1177–1197. doi:10.1111/j.1742-4658.2012.08529.x

31. Vijaya Kumar A, Salem Gassar E, Spillmann D, et al. HS3ST2 modulates breast cancer cell invasiveness via MAP kinase- and Tcf4 (Tcf7l2)-dependent regulation of protease and cadherin expression. *Int J Cancer*. 2014;135(11):2579–2592. doi:10.1002/ijc.28921

32. Dankova Z, Brany D, Dvorska D, et al. Methylation status of KLF4 and HS3ST2 genes as predictors of endometrial cancer and hyperplastic endometrial lesions. *Int J Mol Med*. 2018;42(6):3318–3328. doi:10.3892/ijmm.2018.3872

33. Zuo Q, Zheng W, Zhang J, et al. Methylation in the promoters of HS3ST2 and CCNA1 genes is associated with cervical cancer in Uygur women in Xinjiang. *Int J Biol Markers*. 2014;29(4):e354–362. doi:10.5301/jbm.5000107

34. Hellec C, Delos M, Carpentier M, Denys A, Allain F. The heparan sulfate 3-O-sulfotransferases (HS3ST) 2, 3B and 4 enhance proliferation and survival in breast cancer MDA-MB-231 cells. *PLoS One*. 2018;13(3):e0194676. doi:10.1371/journal.pone.0194676

35. Hwang JA, Kim Y, Hong SH, et al. Epigenetic inactivation of heparan sulfate (glucosamine) 3-O-sulfotransferase 2 in lung cancer and its role in tumorigenesis. *PLoS One*. 2013;8(11):e79634. doi:10.1371/journal.pone.0079634

36. Eyvazi S, Khamaneh AM, Tarhriz V, et al. CpG islands methylation analysis of CDH11, EphA5, and HS3ST2 genes in gastric adenocarcinoma patients. *J Gastrointest Cancer*. 2020;51(2):579–583. doi:10.1007/s12029-019-00290-1